



## The entropy of scale-space

Sporring, Jon

*Published in:*  
Proceedings of 13'th International Conference on Pattern Recognition (ICPR'96)

*Publication date:*  
1996

*Document version*  
Early version, also known as pre-print

*Citation for published version (APA):*  
Sporring, J. (1996). The entropy of scale-space. In *Proceedings of 13'th International Conference on Pattern Recognition (ICPR'96)*

# The Entropy of Scale-Space

Jon Sparring

(sparring@diku.dk)

Department of Computer Science / University of Copenhagen  
Universitetsparken 1 / DK-2100 Copenhagen East  
DENMARK

## Abstract

*Viewing images as distributions of light quanta enables an information theoretic study of image structures on different scales. This article combines Shannon's entropy and Witkin and Koenderink's Scale-Space to establish a precise connection between the Heat Equation and the Thermodynamic Entropy in Scale-Space. Experimentally the entropy function is used to study global textures.*

## 1. Introduction

A recent paper in by Jägersand [3] investigated the so-called Kullback measure as a measure of the change of information in Scale-Space on images. This measure does not perform well in the limit of infinitesimal change as is shown in appendix A. A similar description is based on the derivative of the entropy w.r.t. the natural parameter in Scale-Space. In the present paper, this measure will be derived and applied to the study of textures.

Further and most importantly, the connection between the Heat Equation and the Thermodynamic Entropy is established in the context of Gaussian Scale-Space.

## 2. Discretizing Distributions

The mean entropy [7] of a discretely defined probability distribution  $p$  is defined as,

$$\langle S(p) \rangle = - \sum_{\mathbf{x}} p(\mathbf{x}) \log p(\mathbf{x})$$

which is usually interpreted as the mean information content of the distribution. While it is intuitively easy to interpret entropies of discrete distributions, the same is not true for continuously defined distributions. This paper analyses continuously defined distributions by discretization. Such a

discretization can be performed as,

$$p_{\Delta\mathbf{x}}(i) = \Delta\mathbf{x} \langle p(x = i\Delta\mathbf{x}) \rangle$$

where  $i$  is an integer,  $\Delta\mathbf{x}$  is the discretization constant, and  $\langle p(\mathbf{x}) \rangle$  is the mean value of  $p(\mathbf{x})$  in the  $x \pm \frac{\Delta\mathbf{x}}{2}$  interval. This corresponds to sampling  $p$  convolved with a box filter at points spaced according to  $\Delta\mathbf{x}$ , where the width of the box filter also is  $\Delta\mathbf{x}$ . Many arguments from various fields such as Numerical Analysis [10, p.606ff], Statistical Inference [4], and Physics [6] indicate that the least committed choice of a sampling function is not the box filter, but the Gaussian filter, i.e.

$$p_t(i) = \Delta\mathbf{x}_t (G_t * p) |_{x=i\Delta\mathbf{x}_t}$$

where the sampling factor now is proportional to the standard deviation of the Gaussian function, i.e.  $\Delta\mathbf{x}_t = c\sqrt{t}$ , where  $c$  is an arbitrary constant. The entropy thus becomes,

$$\begin{aligned} \langle S_t(p) \rangle &= - \sum_{\mathbf{i}} \Delta\mathbf{x}_t p_t(i) \log (\Delta\mathbf{x}_t p_t(i)) \\ &= - \log \Delta\mathbf{x}_t - \sum_{\mathbf{i}} \Delta\mathbf{x}_t p_t(i) \log p_t(i) \end{aligned}$$

The last term can be approximated as an integral for easier study by taking an averaging of each interval  $i$ . Thus,

$$\langle S_t(p) \rangle \simeq - \log \Delta\mathbf{x}_t - \int_{\mathbf{x}} p_t(\mathbf{x}) \log p_t(\mathbf{x}) d\mathbf{x}$$

## 3. A Spatial Distribution of Image Points

Think of an image as a realisation of spatially distributed light quanta. Each point in the image registers the number of light quantum-hit. Equivalently, each point is the (un-normalised) probability of light. In this way we may define,

$$p(\mathbf{x}) = \frac{I(\mathbf{x})}{\sum_{\mathbf{x}} I(\mathbf{x})}$$

This spatial probability perspective corresponds very nicely with the theory of Scale-Space [13, 6]. At coarse resolution, the spatial smoothing is high or equivalently, the spatial uncertainty is high and the spatial probability distribution close to being uniform. What further is shown in appendix B is that the spatial entropy complies with the molecular entropy of thermodynamic. This is of importance since The Gaussian Scale-Space can be viewed as a physical model of sampling, as Heat-Diffusion, and thus the study of the Thermodynamic Entropy is one important aspect of this model.

The above distribution is completely invariant to multiplicative constants as a consequence of the normalization, i.e.

$$\frac{kI(\mathbf{x})}{\sum_{\mathbf{x}} kI(\mathbf{x})} = \frac{I(\mathbf{x})}{\sum_{\mathbf{x}} I(\mathbf{x})}$$

However, it is not invariant to additive constants. In brief, the change in entropy when adding a constant to the images,  $I(\mathbf{x}) = I'(\mathbf{x}) + c$  is,

$$\frac{\partial S(I)}{\partial c} = -\frac{1}{\langle I \rangle} (S + \langle \log I \rangle)$$

where  $\langle \bullet \rangle$  is the mean value operator. This function is strictly positive for discrete images, since  $S$  can be shown to be less than  $-\langle \log I \rangle$ , see e.g. Rissanen [11]. An interpretation of this is that: an additive constants changes the function proportional with the difference of the mean of  $\log I$  under the  $I$  distribution and under the uniform distribution.

#### 4. Entropy Change: A characteristic function

As shown in appendix B the entropy as a function of scale is a monotonically growing function, starting with the entropy at scale ‘zero’ and ending at maximum entropy when the mean value image has been reached. The characteristic functionality is thus in the derivatives of the entropy with respect to the natural scale parameter  $\tau = \log t = 2 \log \sigma$ . The natural scale parameter is defined in Florack [6, 2]. The entropy change is thus,

$$\begin{aligned} \frac{d\langle S_t \rangle}{d\tau} &= - \int_{-\infty}^{\infty} (1 + \log p_t(\mathbf{x})) \frac{\partial p_t(\mathbf{x})}{\partial \tau} d\mathbf{x} - \frac{1}{2} \\ &= -t \int_{-\infty}^{\infty} (1 + \log p_t(\mathbf{x})) \frac{\partial p_t(\mathbf{x})}{\partial t} d\mathbf{x} - \frac{1}{2} \end{aligned}$$

It can be seen that the spatial quantization has only constant effect on the information change and will hereafter be ignored.

Using the image as a distribution,  $p_t = \frac{1}{c} I_t = \frac{G_t * I}{c}$ , where  $c = \int I(\mathbf{x}) d\mathbf{x}$ , and  $\frac{\partial I}{\partial t} = 2\nabla^2 I$  (The Heat Equation), where  $\nabla^2$  is the Laplacian operator, thus yields,

$$\frac{d\langle S_t \rangle}{d\tau} = -\frac{2t}{c} \int_{-\infty}^{\infty} (k + \log I_t(\mathbf{x})) \nabla^2 I_t(\mathbf{x}) d\mathbf{x} - \frac{1}{2}$$

where  $k = 1 - \log c$ . The spatial information change is tightly coupled to the second derivatives, as should be expected when using Scale-Space. Numerically, the second order derivatives can only be evaluated at scales higher than the inner scale ( $\sigma \geq 1$ ). On the other hand, experiments have shown that significant change can take place for lower scales. Luckily experiments also seem to indicate that the entropy as a function of scale is a smooth function (for large images), and thus a numerical derivation of the entropy function can be used on lower scales instead.

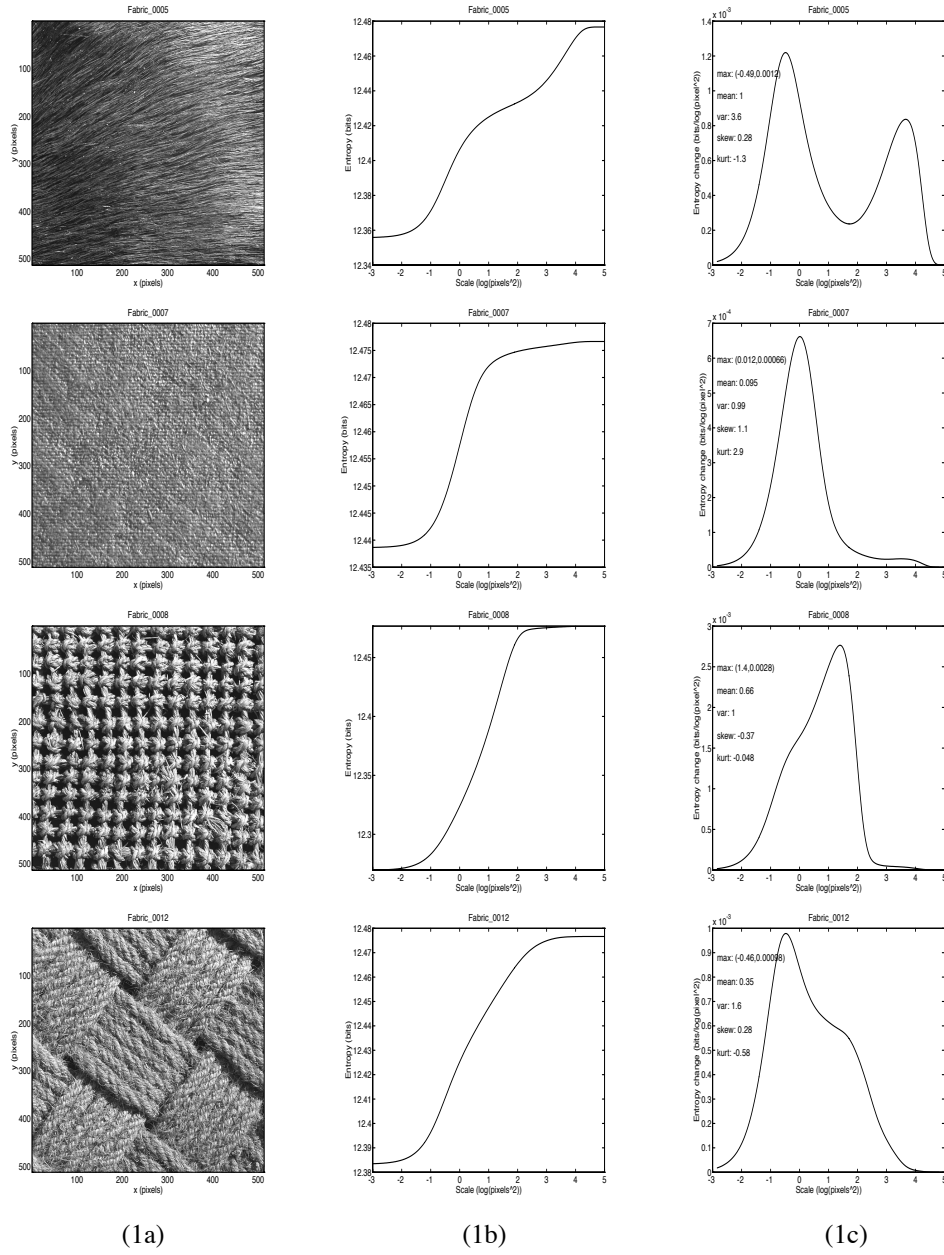
#### 5. Experiments

The entropy change by scale is a global measure, and thus the experiments performed concentrate on characterization of globally defined texture-images. Furthermore, Gauss-convolution has been implemented using Fast Fourier Transformation and therefore the image is implicitly assumed to be on a torus. This will give some ‘peculiar’ boundary effects. This boundary will increase in size with scale while the information contents will decrease. For large images this is not expected to have a dominating effect on the result. All images used in these experiments are from the ‘VisTex’ collection [9].

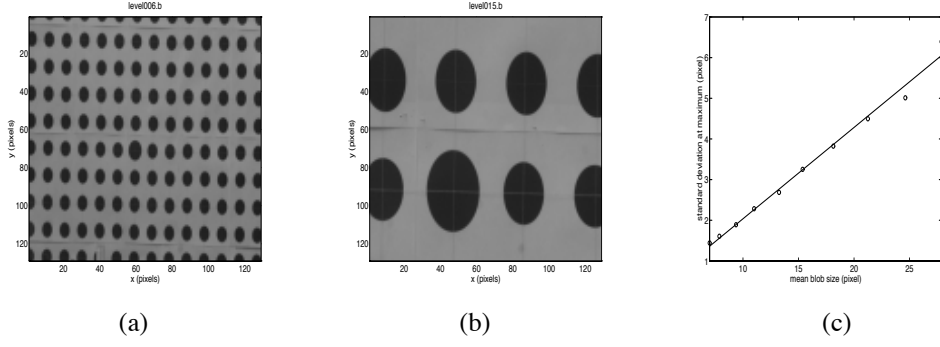
Figure 1 shows examples of the non-resampled entropy functions, i.e. the  $\log \Delta \mathbf{x}_t$  has been ignored. Note the monotonicity of the entropy function, and the regularity of the first derivative with respect to logarithmic scale.

A first order effect to expect from simple textures is that the point of maximal entropy change should correspond to the size of the dominating image structures. Figure 2a-b shows images from a lab-experiment: The camera is placed fronto-parallel to a plane with a very simple texture. A sequence is produced as a series of increasing zoom values. In figure 2c the standard deviation of the point of maximum entropy change has been plotted against the mean size of the small blob’s. As can be seen, the relation is linear except in the last few images. There the larger blob starts to have an effect on the maximum. The  $l_2$  distance error from linearity is 0.4854 to the shown line.

The higher order moments of the entropy change function does not monotonically grow with increased inner scale. The main reason for this being that different structures are dominating at different inner scales. But comparing the entropy change functions in figure 1 with 2 it is seen that the difference in the finer structure results in difference of the functions ‘skew’. It is the experience of this author based on textures from VisTex, that the entropy change function is indeed a characteristic function of a particular texture.



**Figure 1. Examples of the entropy-scale function. The a's are the image/distribution, the b's entropy as a function of logarithmic scale, and the c's the entropy change with respect to the logarithmic scale.**



**Figure 2.** The images in (a) and (b) are the first and the last from a sequence of images of a fronto-parallel plane with circular texture taken with different zoom-value. The graph in (c) is the point of maximum entropy change plotted against the estimated ‘blob’-expansion

## 6. Discussion

This paper has empirically indicated that the derivative of the entropy of the spatial image distribution with respect to the natural scale parameter,  $\log t$ , is characteristic function of textures.

The formal correspondence of the entropy in Scale-Space and the Thermodynamic Entropy has been established. By the second law of thermodynamics follows that the spatial entropy is an increasing function of scale.

One application of the entropy is scale-selection: scales of maximal information loss is the scales of which the dominating image content deteriorates fastest. It might even be possible to globally distinguish several dominating scales. Another application is quantization: Using the entropy change, a spatial down-sampling (a pyramid) can be calculated in such a way, that the information loss is constant. This refines the *a priori* natural logarithmic scale by image content.

A natural extension of this paper would be local entropies, for e.g. texture segmentation. It is of course straight forward to window the images, but this would not be a ‘true’ local method. Another important matter is that the presented entropy function is just one of a continuum of information measuring functions. These two extension are currently being investigated.

## 7. Acknowledgements

To end, I would like to mention that this article would never have come about without the many and very useful insights by Mads Nielsen and Luc Florack. Finally should Niels Holm Olsen be acknowledged for his help with some of the programming.

## A. The Kullback divergence

The Kullback divergence has been used by Jägersand [3] to measure the change of information when images are viewed on different scales in Scale-Space. The Kullback divergence is given as,

$$\begin{aligned} K[p, q] &= \int p(\mathbf{x}) \log p(\mathbf{x}) - \int p(\mathbf{x}) \log q(\mathbf{x}) d\mathbf{x} \\ &= \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} \end{aligned}$$

where  $p$  and  $q$  are two distributions defined on identical  $N$ -dimensional domains to which  $\mathbf{x}$  belongs. Historically, the Kullback divergence is a measure of the waste of band-width when coding with an incorrect distribution.

In the case of the spatial distributions in Scale-Space,  $p$  and  $q$  belong to a continuous one-parameter family of distributions, and hence in the limit of infinitesimal change the change in Kullback is,

$$\begin{aligned} K[p_a, p_{a+\epsilon}] &= \int p_a(\mathbf{x}) \log \frac{p_a(\mathbf{x})}{p_{a+\epsilon}(\mathbf{x})} d\mathbf{x} \\ &= \epsilon \int p_a(\mathbf{x}) \frac{\log p_a(\mathbf{x}) - \log p_{a+\epsilon}(\mathbf{x})}{\epsilon} d\mathbf{x} \end{aligned}$$

It is now straight forward to see that,

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{K[p_t, p_{t(1+\epsilon)}]}{\epsilon} &= \int p_t(\mathbf{x}) \frac{\partial \log p_t(\mathbf{x})}{\partial t} d\mathbf{x} \\ &= \int \frac{\partial p_t(\mathbf{x})}{\partial t} d\mathbf{x} \end{aligned}$$

Thus the value of  $K$  only depends on the values of  $p_t$  in the limits of the integral. The Kullback measure is therefore not a global measure of the information change in Scale-Space in the infinitesimal limit.

## B. Monotonicity of Entropy by Scale

The spatial entropy is equivalent of the thermodynamic entropy as defined in statistical mechanics. Since both Scale-Space and Spatial Entropy mimics thermodynamic processes, the second law of thermodynamics ensures that the spatial entropy is an increasing function of scale. This formal relation will be established in the following. For a reference on thermodynamics and statistical mechanics see for example Atkins [1] and Kittel [5].

The spatial entropy of an image is given as,

$$\langle S_t(I) \rangle = \sum_{\mathbf{x}} \frac{I_t(\mathbf{x})}{\sum_{\mathbf{x}} I(\mathbf{x})} \log \left( \frac{I_t(\mathbf{x})}{\sum_{\mathbf{x}} I(\mathbf{x})} \right)$$

One of the basic assumptions of statistical mechanics is the indifference of thermodynamics to position of the system under study. The number of different arrangements of a fixed number of subsystems is signified by the multiplicity function,

$$g(\{n_i\}) = \frac{N!}{\prod_i n_i!}$$

where  $n_i$  is the number of particles in subsystem  $i$  and  $N = \sum_i n_i$  is the total number of particles in the thermodynamically closed system under study. The thermodynamic (dimensionless) entropy is given as,

$$S = \log g$$

which can be simplified using Stirling's approximation,

$$x! \simeq \sqrt{2\pi x} x^{x+\frac{1}{2}} \exp(-x + \dots)$$

where the higher order terms can be neglected for  $x$  greater than about 10. The thermodynamic entropy is thus approximated as,

$$\begin{aligned} S &= \log g \\ &= \log N! - \sum_i \log n_i! \\ &= \frac{1-k}{2} \log 2\pi + (N + \frac{1}{2}) \log N - N \\ &\quad - \sum_{i=1}^k (n_i + \frac{1}{2}) \log n_i + \sum_{i=1}^k n_i \\ &= \frac{1-k}{2} \log 2\pi + (N + \frac{1}{2}) \log N \\ &\quad - \sum_{i=1}^k (n_i + \frac{1}{2}) \log \frac{n_i}{N} - (N + \frac{k}{2}) \log N \\ &= \frac{1-k}{2} \log 2\pi N - \sum_{i=1}^k (n_i + \frac{1}{2}) \log \frac{n_i}{N} \\ &= c - N < S(N) > - \frac{1}{2} \sum_{i=1}^k \log \frac{n_i}{N} \end{aligned}$$

For large  $N$ 's, the spatial entropy part will be dominating, or in other words, when  $n_i \gg \frac{1}{2}$ , this simplifies to,

$$S \simeq c - N < S(N) >$$

The second law of thermodynamics states that the entropy of a closed system will grow towards equilibrium monotonically with time. Scale-Space is governed by the heat diffusion of a perfect gas, where each intensity value signifies the number of molecules in each spatial position (subsystem). Hence the spatial entropy will grow monotonically with time.

## References

- [1] P. W. Atkins. *Physical Chemistry*. Oxford University Press, 1990.
- [2] L. Florack, B. ter Haar Romeny, J. Koenderink, and M. Viergever. Linear scale-space. *Journal of Mathematical Imaging and Vision*, 4:325–351, 1994.
- [3] M. Jägersand. Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach. In *Fifth International Conference on Computer Vision*, pages 195–202. IEEE Computer Society Press, June 1995.
- [4] E. T. Jaynes. Prior probabilities. *IEEE Transactions on systems science and cybernetics*, 4(3):227–241, 1968.
- [5] C. Kittel and H. Kroemer. *Thermal Physics*. W. H. Freeman and Company, New York, 1980.
- [6] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [7] M. Li and P. Vitányi. *An introduction to Kolmogorov complexity and its applications*. Springer-Verlag, 1993.
- [8] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, The Netherlands, 1994.
- [9] R. Picard, C. Graczyk, S. Mann, J. Wachman, L. Picard, and L. Campbell. Vistex. via ftp:whitechapel.media.mit.edu, 1995. Copyright 1995 Massachusetts Institute of Technology.
- [10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C*. Cambridge University Press, 1992.
- [11] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific, 1989.
- [12] B. M. ter Haar Romeny (Ed.), editor. *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, The Netherlands, 1994.
- [13] A. P. Witkin. Scale space filtering. In *Proc. of International Joint Conference on Artificial Intelligence (IJCAI)*, Karlsruhe, Germany, 1983.